(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) Publication number: **0 635 978 A1**

(12) **EUROPEAN PATENT APPLICATION published in accordance with Art. 158(3) EPC**

(21) Application number: **94902095.2**

(51) Int. Cl.[6]: **H04N 7/01**

(22) Date of filing: **09.12.93**

(86) International application number:
PCT/JP93/01786

(87) International publication number:
WO 94/14278 (23.06.94 94/14)

(30) Priority: **10.12.92 JP 3305/92**
14.06.93 JP 167518/93
18.06.93 JP 172617/93

(43) Date of publication of application:
25.01.95 Bulletin 95/04

(84) Designated Contracting States:
DE FR GB IT NL

(71) Applicant: **SONY CORPORATION**
7-35 Kitashinagawa 6-chome
Shinagawa-ku
Tokyo 141 (JP)

(72) Inventor: **KONDO, Tetsujiro Sony Corporation**
7-35, Kitashinagawa 6-chome
Shinagawa-ku
Tokyo 141 (JP)

Inventor: **UCHIDA, Masashi**
Sony Corporation
7-35, Kitashinagawa 6-chome
Shinagawa-ku
Tokyo 141 (JP)

Inventor: **KAWAGUCHI, Kunio Sony Corporation**
7-35, Kitashinagawa 6-chome
Shinagawa-ku
Tokyo 141 (JP)

(74) Representative: **Ayers, Martyn Lewis Stanley et al**
J.A. KEMP & CO.
14 South Square
Gray's Inn
London WC1R 5LX (GB)

(54) **IMAGE SIGNAL CONVERTER.**

(57) A digitally inputted image signal (SD signal) is converted into a high resolution digital image signal (HD signal) by using estimated values. The specific pixels of an object to be estimated are classified according to the one-, two-, or three-dimension level distribution of plural reference pixels of SD signal located near the specific pixels. The estimated values of the specific pixels are generated by a linear combination of the values of the reference values and the estimation coefficients determined in advance by learning. During the learning, a known HD signal and the SD signal formed from it are used to determine the estimation coefficients so as to minimize the sum of squares of the errors between the true values and the values estimated by the linear combination of the values of the surrounding SD signal reference pixels and the estimated coefficients. Not necessarily limited to the estimated coefficients, it may be possible to use, as the estimated values, representatives determined for every class in correspondence with the class of the inputted SD signal. The reference value of a block and the value normalized by the dynamic range DR are an example as the representatives.

EP 0 635 978 A1

Technical Field

The present invention relates to an image signal converting apparatus suitable for so-called up-conversion where an image signal with a standard resolution is converted into an image signal with a high resolution.

Background Art

For example, in the case of TV signals, in addition to standard resolution (or standard definition SD) TV signals, high definition (or high definition HD) TV signals have been experimentally used for broadcasting. TV receivers that can receive HD signals have been practically used. In other words, the HD TV system is going to coexist with the SD TV system.

When the SD TV system and the HD TV system coexist, a signal converting apparatus for allowing an HD TV monitor to reproduce SD TV video signals will be required. Although various standards for the SD signals and the HD signals have been proposed, now assume an NTSC system for SD signals (number of scanning lines = 525, number of fields = 60, aspect ratio = 4 : 3) and an HDTV system for HD signals (number of scanning lines = 1125, number of fields = 60, aspect ratio = 16 : 9). The resolution of the HD signals is four times higher than the resolution of the SD signals. Thus, the converting apparatus should perform an up-converting process that increases the number of pixels of the input SD signal four times so as to obtain the resolution of an HD signal.

Conventionally, the signal converting apparatus uses interpolating filters. Fig. 1 shows a block diagram of a conventional signal up-converting apparatus. An SD signal is supplied from an input terminal 1. The SD signal is sent to a horizontal interpolating filter 2 that increases the number of pixels of the SD signal twice in horizontal direction. The output of the horizontal interpolating filter 2 is sent to a vertical interpolating filter 3 that increases the number of pixels of the output signal of the horizontal interpolating filter 2 twice in vertical direction. The output of the vertical interpolating filter 3 is an HD signal. The HD signal is sent to an output terminal 4. In the conventional signal converting apparatus, the image is up-converted with such filters. For example, 0 data is inserted into an interpolating point so that the sampling frequency of the SD signal accords with the sampling frequency of the HD signal. An interpolated value corresponding to the interpolating point is formed by each filter.

Fig. 2 shows a construction of each of the interpolating filters. An SD signal is supplied to an input terminal 5. The SD signal is sent to multipliers that multiply the SD signal by filter coefficients $\alpha_n, \alpha_{n-1}, \ldots \alpha_0$. Each of outputs of the multipliers is sent to a register with a unit delay amount T. The output of each of the multipliers and the output of each of the registers T are added. An interpolated output is sent to an output terminal 6. In the horizontal interpolating filter 2, the unit delay amount T accords with a sample period. In the vertical interpolating filter 3, the unit delay amount T accords with a line period.

In the conventional image signal converting apparatus, since the SD signal is up-converted into the HD signal with the filters, the output HD signal is generated by processing the input SD signal. Thus, the resolution of the output HD signal is not higher than the resolution of the input SD signal.

An object of the present invention is to provide an image signal converting apparatus that can compensate the resolution of the output signal.

Another object of the present invention is to provide an image signal converting apparatus that categorizes the input signal as classes corresponding to local features of the input image signal so as to improve conversion accuracy.

Disclosure of Invention

A first aspect of the present invention is a signal converting apparatus for converting a first digital image signal into a second digital image signal having higher resolution than the first digital image signal, comprising a class categorizing means for categorizing a considered pixel of the second digital image signal as a class corresponding to a pattern of level distribution of a plurality of reference pixels of the first digital image signal, the reference pixels of the first digital image signal being disposed spatially and/or chronologically adjacent to the considered pixel, a memory means for storing predicted coefficients for each class and for outputting the predicted coefficients corresponding to class information received from the class categorizing means, and a predicted value generating means for calculating linear combination of the values of the pixels of the first digital image signal spatially and/or chronologically adjacent to the considered pixel so as to generate a predicted value of the considered pixel. The predicted coefficients are pre-learnt and pre-determined.

To compress data, various techniques such as DCT (Discrete Cosine Transform), VQ (Vector Quantizing), DPCM (Differential Pulse Code Modulation), BCT (Block Truncation Coding), and non-linear quantizing may be selectively used as well as the ADRC (Adaptive Dynamic Range Coding) technique.

As shown in Fig. 4, the block segmenting circuit 12 converts raster scanning sequence data of the TV signal into two-dimension block sequence data. In the example shown in Fig. 4, one block is composed of (3 x 3) SD pixels denoted by a to i. The predicted value generating circuit 16 generates the values of four HD pixels A to D nearly at the center of the block corresponding to the SD pixels of the block.

Instead of the block shown in Fig. 4, the predicted value generating circuit 16 may generate predicted values of HD pixels corresponding to a one-dimensional block of four SD pixels a to d as shown in Fig. 5. Moreover, the present invention may be applied to a three-dimensional block that will be descried later.

In the one-dimensional arrangement shown in Fig. 5, the HD pixel A is categorized as a class and a predicted value thereof is generated corresponding to the SD pixels a, b, and c. The HD pixel B is categorized as a class and a predicted value thereof is generated corresponding to the SD pixels a, b, c, and d. Likewise, these class categorizing process and a predicted value generating process apply to the learning operation.

Next, the ADRC technique used for the data compressing circuit 13 will be described. In the ADRC technique, redundancy of level is adaptively removed corresponding to the correlation of local features of pixels. As shown in Fig. 6, it is clear that dynamic ranges A and B in each block that is re-quantized are much smaller than the original dynamic range of 0 to 255 of the original data composed of eight bits. Thus, the number of bits necessary for re-quantizing these blocks is much smaller than the eight bits.

Now, assume that the number of assigned bits that is smaller than the number of original bits (eight bits) is p, the dynamic range of a block is DR, the value of a pixel in the block is x, and a re-quantized code is Q. With formula (1), the level between maximum value MAX and minimum value MIN is equally divided into 2^p portions and then re-quantized. Fig. 7(a) shows a re-quantizing operation of a level where p is 3.

$$DR = MAX - MIN + 1$$
$$Q = [(x - MIN + 0.5) \times 2^p / DR] \quad (1)$$

where [z] represents the maximum integer that is z or less.

Next, the mean value of pixels in the block having the data level equivalent to $(2^p - 1)$ in the p-bit re-quantizing graduation level in Fig. 7(a) is calculated. The calculated result is a new maximum value MAX'. The mean value of the pixels in the block having the data level equivalent to the re-quantizing graduation level 0 is a new minimum value MIN'. With the new maximum value MAX' and the new minimum value MIN', the dynamic range is defined. With formula (2), the re-quantizing operation is performed.

$$DR' = MAX' - MIN'$$
$$q = [(x - MIN') \times (2^p - 1) / DR' + 0.5] \quad (2)$$

where [z] represents the maximum integer that is z or less.

In the ADRC technique where the new maximum value MAX', the minimum value MIN', and the dynamic range DR' are re-defined, information amount can be effectively compressed. The compressed result is not affected by noise.

As a quantizing operation corresponding to the ADRC technique, representative level being restored may have the same level as the maximum value MAX and the minimum value MIN.

As an example of a two-dimensional block, corresponding to the above-described ADRC technique, values of n pixels where eight bits with values a to i have been compressed to p bits are sent to the class code generating circuit 14. Thus, a code class representing class c is generated corresponding to formula (3).

$$class = \sum_{i=1}^{n} q_i (2^p)^i \quad (3)$$

Next, with reference to Fig. 8, an ADRC circuit corresponding to the one-bit ADRC technique of will be described. In Fig. 8, a block sequence data is supplied to an input terminal 21. The block sequence data is sent to a detecting circuit 22 that detects the maximum value MAX and the minimum value MIN of pixels for each block. The MAX and MIN are sent to a subtracting circuit 23. The output of the subtracting circuit

The horizontal thin-out filter 43 thins out the HD signal in horizontal direction so as to reduce the number of pixels of the HD signal in horizontal direction by 1/2. Thus, an image signal with pixels same as an SD signal is sent to a learning portion 44. A predicted coefficient memory 45 stores predicted coefficients w1 to wn at an address corresponding to the class determined by a class categorizing circuit in the learning portion 44.

The method for forming an SD signal from an HD signal is not limited to the above-described method where the thin-out filters are used. Instead, as in the pixel arrangement shown in Fig. 4, the value of an SD pixel a may be formed by the mean value of (2 x 2) pixels (for example A, B, C, and D). Alternatively, unlike with such a simple mean value, a weight mean value of HD pixels in a wide range (for example, HD pixels in one block) may be used for an SD pixel.

As in the arrangement shown in Fig. 4, when one block is composed of (3 x 3) SD pixels, the SD pixels a to i and the HD pixels A, B, C, and D become one set of learnt data. When there are many sets of learnt data for one frame and the number of frames is increased, a large number of sets of learnt data can be used.

The learning portion 44 performs both a class categorizing process and a calculating process. As the class categorizing process, the learning portion 44 compresses the SD pixels a to i and categorizes the pattern of the two-dimensional distribution of the values of the compressed SD pixels as a class. As the calculating process, the learning portion 44 determines predicted coefficients of each class corresponding to the method of least squares. The class categorizing process performed by the learning portion 44 is the same as the process performed by the data compressing circuit 13 and the class code generating circuit 14 shown in Fig. 3. This learning portion 44 may be accomplished by software. Fig. 12 is a flow chart showing the process of software for the learning portion 44.

At step 51, the control of the learning portion 44 is started. At step 52 "SEGMENT CORRESPONDING DATA INTO BLOCK", an HD signal and an SD signal are supplied and a process for extracting the HD pixels (A to D) and the SD pixels (a to i) in the arrangement shown in Fig. 4 is performed. At step 53 "DATA END ?", if data process for all data (for example, one frame) has been completed, the flow advances to step 56 "DETERMINE PREDICTED COEFFICIENTS". Otherwise, the flow advances to step 54 "DETERMINE CLASS".

At step 54 "DETERMINE CLASS", a class is determined corresponding to the pattern of the level distribution of the SD pixels (a to i) in the vicinity of the HD pixels (A to D of Fig. 4) to be predicted. In this process, as described above, so as to reduce the number of bits, the SD pixels are compressed by for example the ADRC technique. At step 55 "GENERATE NORMAL EQUATIONS", formulas (12), (13), and (14), which will be descried later, are generated.

At step 53 "DATA END ?", it is determined whether or not all data has been completely processed. When all data has been completely processed, the flow advances to step 56. At step 56 "DETERMINE PREDICTED COEFFICIENTS", formula (14), which will be described later, is solved by matrix solution so as to determine predicted coefficients. At step 57 "STORE PREDICTED COEFFICIENTS", the predicted coefficients are stored in the memory. At step 58 "END", the sequence of the steps of the process of the learning portion 44 is completed.

Like the above-described signal converting apparatus, with the pixel arrangement shown in Fig. 4, the SD pixels a to i are compressed and encoded. The encoded values are categorized as a class. Predicted values of the HD pixels A to D are generated by linear combination of the values of the SD pixels a to i and the predicted coefficients w_1 to W_n with a combination shown in Fig. 9.

Next, the process for extracting predicted coefficients for HD pixels from the values of SD pixels will be described in detail. Assume that the values of SD pixels are x_1 to x_n and the true value of the considered HD pixel is y. For each class, a linear combination of n taps (coefficients w_1 to w_n) is given as follows.

$$y' = w_1x_1 + w_2x_2 + ... + w_nx_n \quad (7)$$

In formula (7), before the learning process is performed, w_i is an unknown coefficient.

As described above, the learning process is performed for a plurality of HD data and a plurality of SD data for each class. When the number of data is m, formula (7) is modified to formula (8).

$$y_j' = w_1x_{j1} + w_2x_{j2} + ... + w_nx_{jn} \quad (8)$$

where $j = 1, 2, ... , m$.

When $m > n$, since w_1 to w_n are not unique values, elements of an error vector are defined as follows.

In the conventional signal converting apparatus, to use moving information, the motion of an image is detected. Moving pixels are processed in a field, whereas still pixels are processed among fields. The output signal of each process is mixed with a coefficient corresponding to the amount of motion. Thus, in the conventional apparatus, a motion detecting circuit is additionally required. In addition, unless the motion is accurately detected, image deterioration tends to occur. The class categorizing process and the predicted value generating process using SD pixels that are three-dimensionally arranged can solve such problems.

As described above, the class categorizing process and the predicted value generating process for the signal converting apparatus, which converts an SD signal into an HD signal, are the same as those processes for the learning portion. Next, the class categorizing process and the predicted value genorating process for SD pixels that are three-dimensionally arranged will be described. First, the relation of positions of lines of an HD signal and an SD signal will be explained.

As with the learning process, when an HD signal is converted into an SD signal, pixels on three successive lines of the HD signal are multiplied by non-symmetrical coefficients such as 1/8, 4/8, and 3/8 as shown in Fig. 13, one line of the SD signal is formed. These non-symmetrical coefficients are used in the reverse order for each field so as to maintain the interlace structure of the SD signal. In other words, if two lines of the HD signal are combined and one line of an SD image is formed, since the intervals of lines in the n-th field do not accord with the intervals of lines in the (n + 1)-th field, the interlace structure will be lost.

For the class categorizing process and the predicted value generating process, for example 12 SD pixels that are three-dimensionally distributed are used. There are four patterns of 12 SD pixels corresponding to the position of one HD pixel to be predicted. These four patterns are referred to as mode 1, mode 2, mode 3, and mode 4. Figs. 14, 15, 16, and 17 show arrangements of pixels corresponding to mode 1, mode 2, mode 3, and mode 4, respectively. In Figs. 14 to Fig. 17, circles represent HD pixels to be predicted and squares represent SD pixels to be used for class categorizing process and predicted value generating process.

For the simplicity of hardware, in Fig. 15 and 17, SD pixels in the (n - 1)-th field and the (n + 1)-th field are partially averaged so as to form SD pixels in the n-th field. The SD pixels in the n field are denoted by squares with dotted lines. Figs. 14, 15, 16, and 17 each show a plurality of SD pixels and one HD pixel where the SD pixels are used for the class categorizing process and the predicted value generating process and the HD pixel is to be predicted.

In mode 1 (arrangement shown in Fig. 14), six SD pixels in the n-th field, three SD pixels in the (n - 1)-th field, and three SD pixels in the (n + 1)-th field (a total of 12 SD pixels) are used to perform the class categorizing process and the predicted value generating process for an HD pixel. In mode 1, an HD pixel on the line where a coefficient of 1/8 is multiplied is predicted (see Fig. 13).

In mode 2 (arrangement shown in Fig. 15), SD pixels on the same line in the (n - 1)-th field and the (n + 1)-th field are averaged so as to interpolate three SD pixels that are required in the n-th field. The six SD pixels including the interpolated SD pixels in the n-th field, three SD pixels other than SD pixels used for averaging in the (n - 1)-th field, and three SD pixels other than SD pixels used for averaging in the (n + 1)-th field are used. Thus, a total of 12 SD pixels are used. In mode 2, an HD pixel on the line where a coefficient of 4/8 is multiplied is predicted (see Fig. 13).

In mode 3 (arrangement of Fig. 16), six SD pixels in the n-th field, four SD pixels in the (n - 1)-th field, and two SD pixels in the (n + 1)-th field (a total of 12 SD pixels) are used. In mode 3, an HD pixel on the line where a coefficient of 1/8 is multiplied is predicted (see Fig. 13).

In mode 4 (arrangement of Fig. 17), SD pixels in the (n - 1)-th field and (n + 1)-th field are averaged so as to interpolate four SD pixels (two pixels each on two lines) that are required in the n-th field. The eight SD pixels including the interpolated SD pixels in the n-th field, two SD pixels other than the SD pixels used for averaging in the (n - 1)-th field, and two SD pixels other than the SD pixels used for averaging in the (n + 1)-th field are used. Thus, a total of 12 SD pixels are used. In mode 4, an HD pixel on the line where a coefficient of 4/8 is multiplied is predicted (see Fig. 13).

As described above, in modes 3 and mode 4, an HD pixel is generated at the 1/2 position of the sampling intervals of SD pixels. In other words, in mode 1 to mode 4, an HD pixel with horizontal and vertical resolutions that are twice those of an SD pixel is generated.

In the learning process, predicted coefficients of each of mode 1 to mode 4 are determined by the method of least squares for each class and stored in the memory. Fig. 18 shows a construction of a signal converting apparatus that generates an output HD image corresponding to an input SD image.

A signal d0 that is an SD image in the (n + 1)-th field, is supplied to an input terminal 61. The d0 signal is sent to a field memory 62 and up-converting circuits 64a to 64d. The field memory 62 outputs an SD image signal in n-th field. This signal is denoted by signal d1. The signal d1 is sent to the up-converting

The formula (16) represents the process for successively storing the values of center of gravity obtained in the memory. On the other hand, the formula (17) represents the process for storing the cumulated value of the former values of center of gravity in the memory and for dividing the total of the cumulated value and the present value of center of gravity by the number of times of cumulation. As described above, formulas (16) and (17) represent the process for cumulating values normalized with the base value base and the dynamic range DR rather than the true value hd of the HD pixel. When the cumulated value for each class is stored in the memory and the cumulated value is divided by the number of times of cumulation, representative values may be obtained. However, since the cumulated value becomes large, the number of bits of the memory may increase.

The value of center of gravity that has been determined and learnt is used for the signal converting process. In other words, as with the first embodiment, a class code representing a class is sent to the memory. The class code is stored in the memory as an address. The value of center of gravity g(c) of the class is read from the memory. With the value of center of gravity g(c), predicted value hd' is generated corresponding to formula (18).

$$hd' = DR \times g(c) + base \quad (18)$$

In the first and second embodiments of the present invention, when blocks with dynamic range DR that is less than a predetermined value are excluded from data to be learnt, the learning process can be prevented from being affected by noise.

In addition, according to these embodiments, predicted coefficients and representative values are pre-learnt. However, in a high speed system, the predicted coefficients and representative values can be obtained on real time basis. The predicted coefficients and representative values that have been learnt can be updated with those obtained on the real time basis.

Either the method for predicting the values of all pixels of the output image signal with high resolution or the method for predicating the values of pixels that are not present may be selected.

Moreover, the present invention may be applied to image signals other than interlace scanning TV signals, sequence scanning TV signals.

Claims

1. A signal converting apparatus for converting a first digital image signal into a second digital image signal having higher resolution than the first digital image signal, comprising:

   class categorizing means for categorizing a considered pixel of the second digital image signal as a class corresponding to a pattern of level distribution of a plurality of reference pixels of the first digital image signal, the reference pixels of the first digital image signal being disposed spatially and/or chronologically adjacent to the considered pixel;

   memory means for storing predicted coefficients for each class and for outputting the predicted coefficients corresponding to class information received from said class categorizing means; and

   predicted value generating means for calculating linear combination of the values of the pixels of the first digital image signal spatially and/or chronologically adjacent to the considered pixel so as to generate a predicted value of the considered pixel.

2. The signal converting apparatus as set forth in claim 1, further comprising a learning device for predetermining the predicted coefficients,

   wherein said learning device comprises:

   means for forming the first digital image signal with the second digital image signal;

   class categorizing means for categorizing a considered pixel of the second digital image signal as a class corresponding to a pattern of level distribution of a plurality of reference pixels of the first digital image signal, the reference pixels of the first digital image signal being disposed spatially and/or chronologically adjacent to the considered pixel; and

   means for determining predicted coefficients for each class so that the sum of square of an error between the true value of the considered pixel and the predicted value becomes minimum when the predicted value of the considered pixel is generated by linear combination of a plurality of pixels of the first digital image signal and the predicted coefficients.

3. A signal converting apparatus for converting a first digital image signal into a second digital image signal having higher resolution than the first digital image signal, comprising:

Fig. 1

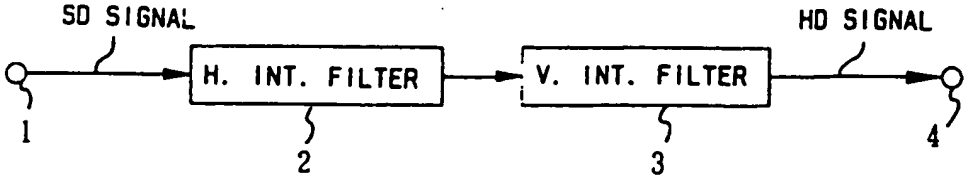


Fig. 2

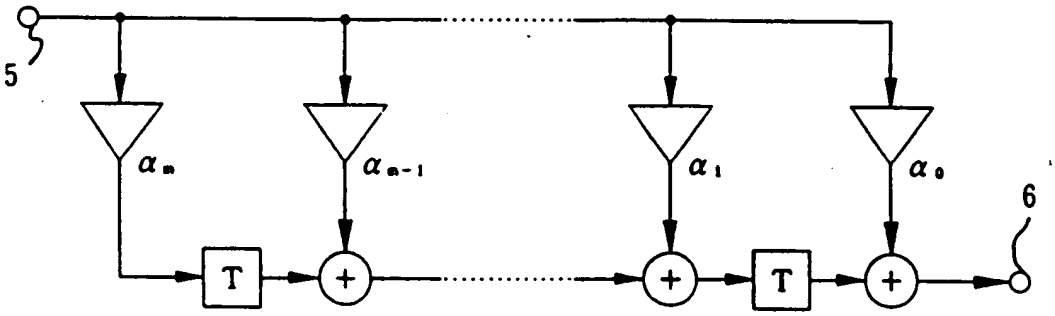


Fig. 3

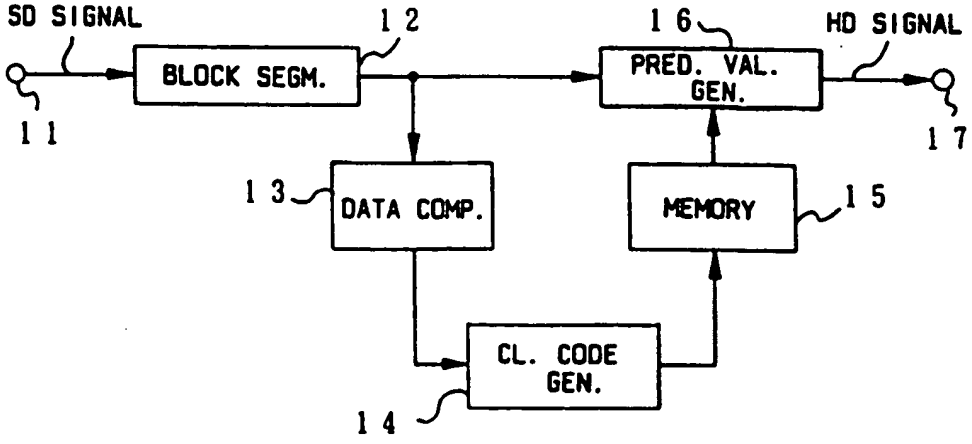


Fig. 6

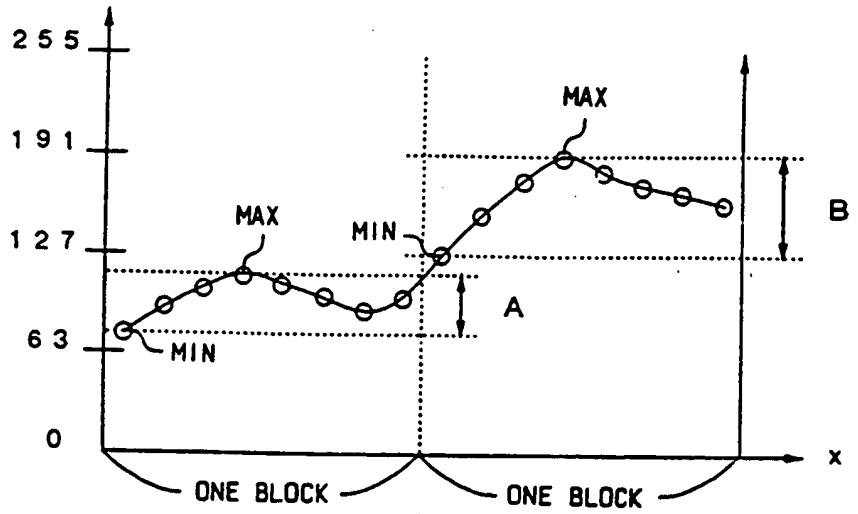

255
191
127
63
0
MAX
MIN
MAX
MIN
A
B
x
ONE BLOCK
ONE BLOCK

Fig. 7

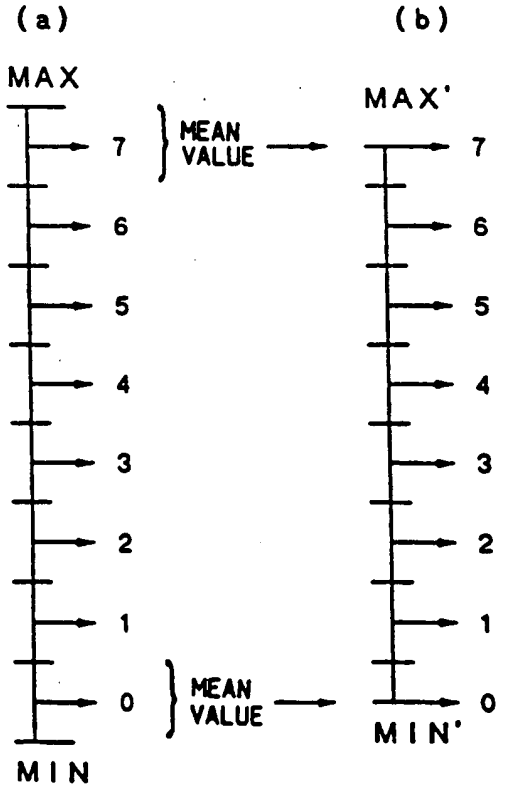

(a)
(b)
MAX
MAX'
7
6
5
4
3
2
1
0
MEAN VALUE
MEAN VALUE
MIN
MIN'

Fig. 10

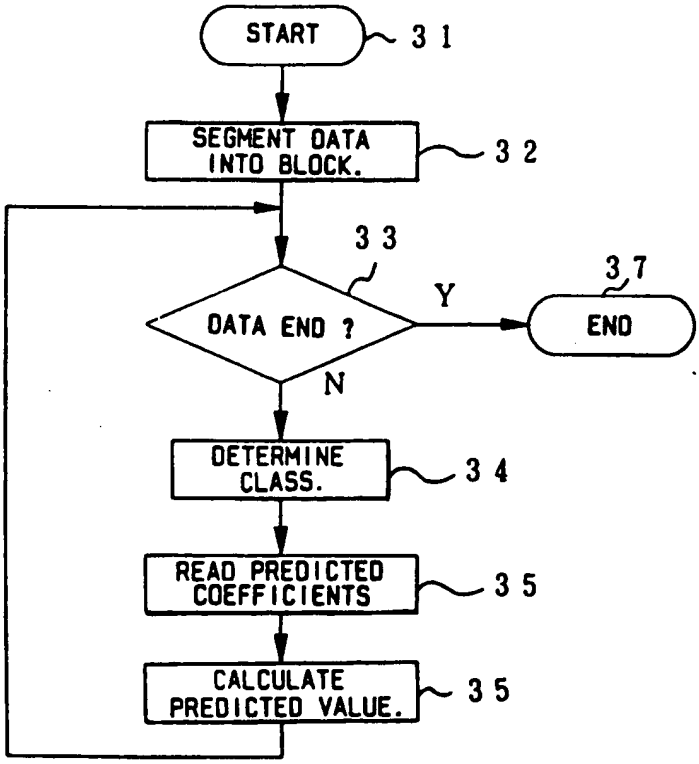

START
31
SEGMENT DATA INTO BLOCK.
32
33
DATA END ?
Y
37
END
N
DETERMINE CLASS.
34
READ PREDICTED COEFFICIENTS
35
CALCULATE PREDICTED VALUE.
35

Fig. 11

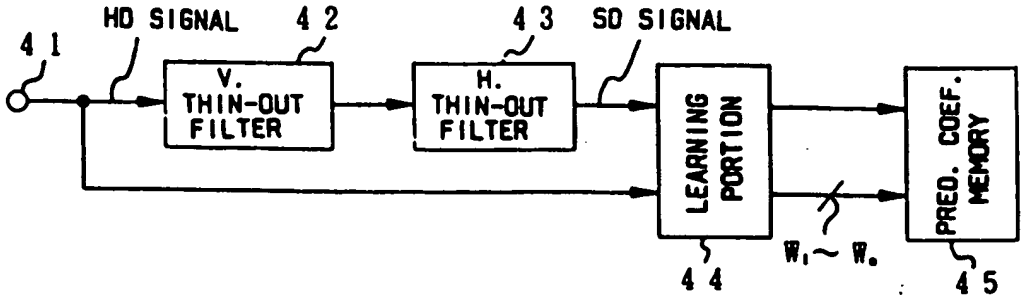

41
HD SIGNAL
42
V. THIN-OUT FILTER
43
H. THIN-OUT FILTER
SD SIGNAL
LEARNING PORTION
44
W1 ~ Wn
PRED. COEF. MEMORY
45

Fig. 13

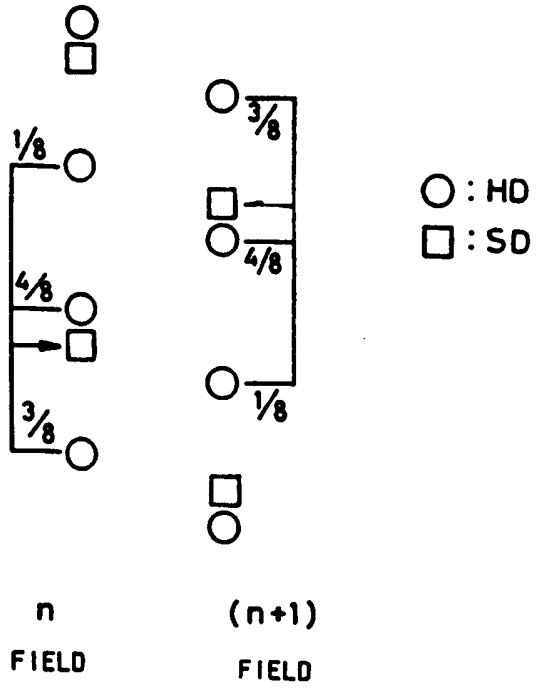
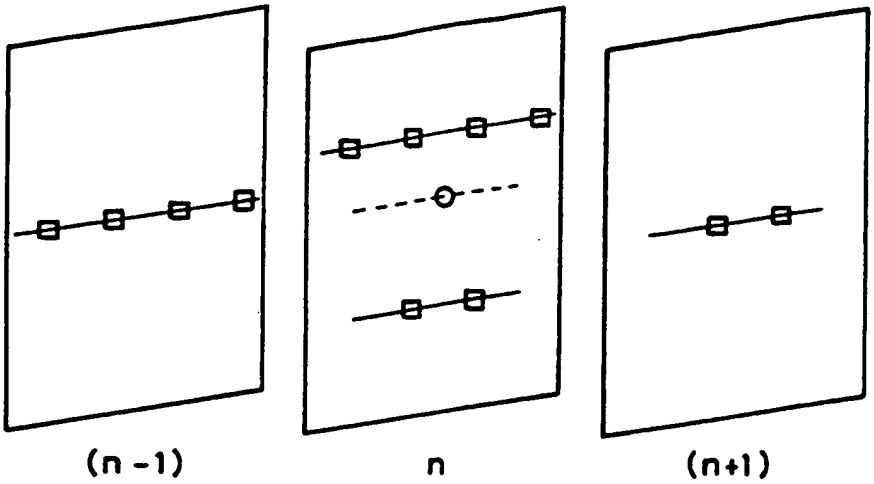
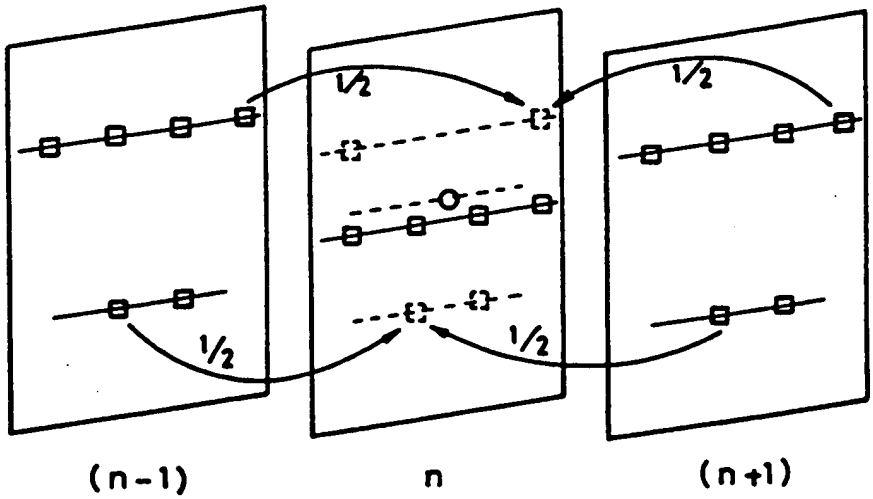


Fig. 16



MODE 3

Fig. 17



MODE 4

12 : BLOCK SEGMENTING CIRCUIT

13 : DATA COMPRESSING CIRCUIT

15 : MEMORY

16 : PREDICTED VALUE GENERATING CIRCUIT

22 : MAXIMUM VALUE/MINIMUM VALUE DETECTING CIRCUIT

44 : LEARNING PORTION